DOCUMENT RESUME

ED 395 985 TM 025 148

AUTHOR Fraas, John W.; Newman, Isadore
TITLE The Importance of Analyzing Longitudinal Data in a Formative Evaluation Process: Applying Statistical Quality Control Techniques.
PUB DATE Feb 96
NOTE 27p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Boston, MA, February 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Data Analysis; *Educational Change; Feedback; *Formative Evaluation; *Longitudinal Studies; Outcomes of Education; Program Evaluation; *Quality Control; Systems Approach; Teacher Evaluation; Teaching Methods; *Total Quality Management
IDENTIFIERS Western Electric

ABSTRACT

One may receive the most benefit from an evaluation of an educational program or the performance of a teacher if the evaluation process is approached from a Total Quality Management (TQM) point of view. Under the philosophy of TQM, the purpose of any evaluation process is to provide feedback for the continual improvement of the educational process that is being evaluated. In the process of obtaining this feedback, the evaluator must be cognizant of two concepts that are basic to the TQM philosophy. First, an educational program or the work of a teacher can be viewed as a system. Second, since every system is impacted by numerous factors, any outcome variable of the system will experience variation. The implication of these two concepts is that an educational evaluator must be able to separate the impact of a change in an educational system from the noise of that system. This paper discusses how an evaluator can increase the ability of an individual values chart to separate a signal in an outcome variable, a signal caused by a change in the educational program or in the teacher's pedagogical practices, from the normal variation found in that variable when the chart is used in conjunction with the Western Electric Zone tests. In an individual values chart, the outcome variable is plotted on a chart in which the time factor and the outcome variable are placed on the X axis and Y axis respectively. The Western Electric Company has proposed four statistical control detection tests to be used in conjunction with control charts to increase the chances of detecting such a signal. (Contains two tables, three figures, and nine references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 395 985

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
JOHN W. FRAAS
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Running head: THE IMPORTANCE OF ANALYZING

The Importance of Analyzing Longitudinal Data in a Formative Evaluation Process: Applying Statistical Quality Control Techniques

John W. Fraas

Ashland University

Isadore Newman

The University of Akron

Paper Presented at the Annual Meeting of The Eastern Education Research Association,

February, 1996, Boston, MA

TM025148

BEST COPY AVAILABLE

Abstract

One may receive the most benefit from an evaluation of an educational program or the performance of a teacher if the evaluation process is approached from a Total Quality Management (TQM) point of view. Under the philosophy of TQM, the purpose of any evaluation process is to provide feedback for the continual improvement of the educational process that is being evaluated. In the process of obtaining this feedback, the evaluator must be cognizant of two concepts that are basic to the TQM philosophy. First, an educational program or the work of a teacher can be viewed as a system. Second, since every system is impacted by numerous factors, any outcome variable of the system will experience variation. The implication of these two concepts is that an educational evaluator must be able to separate the impact of a change in an educational system from the noise of that system. This paper presents how an evaluator can increase the ability of an individual values chart to separate in an outcome variable a signal, which was caused by a change in the educational program or the teacher's pedagogical practices, from the normal variation found in that variable when the chart is used in conjunction with the Western Electric Zone Tests.

The Importance of Analyzing Longitudinal Data in a Formative Evaluation Process: Applying Statistical Quality Control Techniques

Evaluation is conducted to serve either of two purposes. One type of evaluation, which is referred to as summative evaluation, is used to provide information that will be used to judge the effectiveness of an educational program, a faculty member, and/or the performance level of a student. The other type of evaluation, which is referred to as formative evaluation, is undertaken to provide information that can be used to improve the effectiveness of the program, the faculty, or the staff member.

A number of authors, including Deming (1986), Newman and Newman (1993), Langford and Cleary (1995), and Newman and Morsches (1995) have expressed the view that it is essential to think of evaluation as a means of providing feedback that can be used to improve the educational program, the teacher's performance, and ultimately, the student's level of education. Newman and Morsches expressed the view that evaluation is a continual process that monitors and seeks to improve the educational system by adding to it elements that will increase its effectiveness and eliminating from it elements that hinder its effectiveness. Only formative evaluation can serve this purpose.

The Total Quality Management and Evaluation

If one is to implement an effective formal evaluation process of an educational program or a teacher, two concepts must be understood. First, the educational process is systemic. Second, every outcome variable produced by a system contains variation.

Education Viewed as a System

An educational program or a teacher's performance should be viewed as a system or a subsystem. As stated by Langford and Cleary (1995):

> What we see is that a system--in this case, a school system--is made up of a number of subsystems. Each of the subsystems can be defined in the same terms as the larger system: suppliers and inputs; process; outputs organized toward a purposeful end; and, of course, customers (p. 21).

Newman and Morsches (1995) suggest that this systemic view of an educational program or a teacher's performance has an important implication for the evaluation process implemented in an educational setting. Newman and Morsches stated that "when one talks about a systemic conceptualization of a problem, one is actually implying that the problem has to be looked at in context with all the other forces that may influence the problem within that setting" (p. 7).

Thus, in the evaluation process, the influence that these other forces have on the monitored outcome variable or variables must be considered. An evaluator can not identify a possible benefit of a change in an educational program or a change in a teacher's pedagogical practices without measuring the impact of other forces on the system.

Variation is Everywhere

A second important concept for an evaluator to understand is that outcome variables generated by any system will contain variation (Deming, 1986; Juran, 1979; Wheeler, 1993). Regardless of how well a system is designed or managed its outcomes will not be exactly the same because of the changes in the numerous forces that impact that program. Educational programs and a teacher's performance, which can be viewed as systems or subsystems, are no

exceptions. If the major purpose of evaluating is to provide information regarding the impact that a change in the educational program, the faculty member's instructional methods, or the staff member's work procedures has on the outcome variable, the evaluator can not assume that a fluctuation in the outcome variable indicates that the change in the system has caused that fluctuation. Deming (1986) has demonstrated through his funnel simulation that the outcome variable will behave in an erractic fashion when the system is adjusted or modified every time the value of the outcome variable changes.

The evaluator must be able to separate a signal that the change has, indeed, had an impact on the outcome variable from the variation or noise produced by the other factors that affect the outcome variable. The inability of the evaluator to separate a signal from noise can lead to incorrect conclusions being made regarding the effectiveness of the change in the system. If, in deed, the modification in the educational program, teacher's pedagogical practices, or staff member's procedures has had a positive impact on the outcome variable and the evaluator is unable to identity the corresponding signal in the data, the beneficial modification may not be continued. This type of incorrect decision is referred to as a type II error. On the other hand, if a modification does not have a positive impact on the outcome variable but the evaluator incorrectly identifies a signal in the data, the inappropriate modification may be allowed to become part of the system. This type of incorrect decision is referred to as a type I error. An evaluator must be cognizant of the likelihood of committing either of these two types of errors.

An Individual Values Chart: Separating A Signal from Noise

Lanford and Cleary (1995), as well as other authors, have recommended that an individual values chart, which is also referred to as a runs chart, can be used by educational

evaluators to separate a signal from the noise in an outcome variable. In an individual values chart, the outcome variable is plotted on a chart in which the time factor and the outcome variable are placed on the X axis and the Y axis, respectively.

In a book entitled Economic Control of Quality of Manufactured Product, Shewhart (1931) proposed that a technique, which is referred to as the three-sigma rule, be used to separate a signal in the outcome variable from the noise contained in that variable. When using this technique an evaluator would identify a signal in the outcome variable when a value for any given period of time is located more than three standard deviation units above or below the mean of the outcome values in the base period.

Western Electric Zone Tests

Educational evaluators who use an individual values chart to identify whether a change made in the educational system has had an impact on the outcome variable usually employ the three-sigma rule. It is important to understand that the use of three-sigma rule will enable an educational evaluator to detect large changes in the outcome variable. Wheeler (1994) noted that "detection Rule One [the use of three-sigma rule] is ... very sensitive to large shifts (greater than 3.0 SD(X))" (p. 134).

We believe, however, that a modification that is incorporated into the educational system may produce a change in the monitored outcome variable, but the change, which may be educationally important to maintain, will not be large. In such a case it would be important to use a technique that would increase the evaluator's ability to detect a signal when using an individual values chart. The Western Electric Company (Statistical Quality Control Handbook, 1956) proposed that four detection tests be used in conjunction with control charts to increase the

chances of detecting a signal. Wheeler (1994) noted that the use of the Western Electric Zone Tests "improve sensitivity of the control chart to moderate shifts" (p. 134). In his book entitled Advanced Topics in Statistical Process Control, Wheeler provides information regarding the degree to which the sensitivity of the control chart increases for moderate shifts in the outcome variable when the Western Electric Zone Tests are used.

To use these detection rules, an evaluator would calculate one-sigma and two-sigma limits in addition to the three-sigma limits. After the individual values chart is constructed and the one-, two-, and three-sigma limits are calculated and placed on the chart, the evaluator would use the four Western Electric Zone Tests to detect a signal in the outcome variable. A signal would be detected by the four tests when:

Test 1: One value of the outcome variable is located outside of the three-sigma limits.

Test 2: At least two out of three successive values of the outcome variable are located beyond the same two-sigma limit.

Test 3: At least four out of five successive values of the outcome variable are located beyond the same one-sigma limit.

Test 4: Eight successive values of the outcome variable are located on the same side of the mean of the outcome variable.

If the criterion of any of these four tests was meet, the evaluator would declare that a signal was present in the outcome variable. That is, the evaluator would have some indication that the change in the educational system has had impact on the outcome variable. If the criteria of the four detection rules were not met, the evaluator would attribute any change in the outcome variable to noise, that is, the normal variation in the variable. In such a case, the

BEST COPY AVAILABLE

evaluator could not conclude that the chart revealed that the outcome variable changed after the change in the educational system had been implemented.

A Hypothetical Example

How an educational evaluator can apply the Western Electric Zone Tests in conjunction with an individual values chart can best be understood through an example. In the hypothetical example provided in this section, we are assuming that the administration of a university is attempting to determine if a change in the allocation of financial aid has lead to an improvement in the academic levels of the university's first- year students. It was decided that the academic quality of the matriculating students would be measured by the difference between the national mean score of the American College Testing Assessment (ACT) and the mean ACT score of the students who enrolled at the university. The difference between the hypothetical national ACT mean score and the hypothetical ACT mean score of the university's first-year students, which will be referred to as the difference score, was listed in Table 1 for each year of a 12-year period prior to the change in the financial aid allocation procedure.

Insert Table 1 about here

To construct an individual values chart, the evaluator would complete the following steps:

Step 1: The difference scores are plotted on the chart contained in Figure 1. The Y axis is labeled Difference Scores and it is scaled from -1.3 to +.9 . The X axis is labeled Time Period, and it is scaled in increments of one starting with zero.

Insert Figure 1 about here

Step 2: The mean of the difference values ($\bar{x}$)is calculated by dividing the sum of the difference scores by the number of difference scores as follows:

$$\bar{x} = -2.4/12$$

$$\bar{x} = -0.2$$

This mean value is represented on the chart contained in Figure 1 by a solid line.

Step 3: Each moving range value (mR) is calculated by subtracting the smaller difference score from the larger difference score for two successive time periods. Since moving range value can not be calculated for the first time period, the number of moving range values will be one less than the number of time periods. The 11 mR values calculated for the 12 time periods are listed in Table 1.

Step 4: The mean of the moving ranges ($m\bar{R}$) is calculated by dividing the sum of the mR values by the number of mR values as follows:

$$m\bar{R} = 4.1/11$$

$$m\bar{R} = 0.37$$

Step 5: An estimate of the standard deviation value of the difference scores (sigmaX) is calculated by dividing the mR value by 1.128. The 1.128 is a value used to estimate the standard deviation value from the range when the sample size is one, which is the case for individual values charts. The sigmaX value is calculated as follows:

sigmaX = 0.37/1.128

sigmaX = 0.33

Step 6: The one-, two-, and three-sigma limits are calculated as follows:

one-sigma limits $= \bar{x} \pm 1(\text{sigmaX})$

$= \bar{x} \pm 1(.33)$

= -.53 and .13

two-sigma limits $= \bar{x} \pm 2(\text{sigmaX})$

$= \bar{x} \pm 2(.33)$

= -.87 and .47

three-sigma limits $= \bar{x} \pm 3(\text{sigmaX})$

$= \bar{x} \pm 3(.33)$

= -1.19 and .79

Step 7: The four detection tests are applied to the 12 difference scores. If a signal is not detected, the evaluator would consider the system to be stable during the 12 initial time periods. The application of the four detection tests to the individual values chart contained in Figure 1 indicates that the process is stable. If the system had been judged to be stable, the evaluator would need to be cautious of how any signal detected in future data would be interpreted. That is, a signal may not indicate that the outcome variable changed due to the change in the educational system--in this case a change in the financial aid allocation procedure--but rather by some other factor.

Step 8: The evaluator would apply the four detection tests to each successive difference score recorded after the base time period. For the sake of demonstration, two different sets of

difference scores, which are listed in Table 2, have been generated. Each set of scores represents a different scenario.

Insert Table 2 about here

Under Scenario 1, the difference scores that were recorded for the next two years did not exceeded the three-sigma limits. Thus, if the evaluator used only the three-sigma rule, a signal would not be detected. If the evaluator applied the four detection tests, however, a signal would be detected by Test 2 within two years of the time that the change in the system was in place. As Figure 2 reveals, at least two of three successive values were located outside of the same two-sigma limit during the twelfth, thirteenth, and fourteenth time periods. Thus, the evaluator would have some evidence that the outcome variable, that is, the difference scores had changed after the financial aid allocation procedures had been modified.

Insert Figure 2 about here

Again, under Scenario 2, the difference scores that were recorded for the next four years did not exceed the three-sigma limits. The application of the four detection tests, however, would reveal a signal. As revealed in Figure 3, the application of Test 3, which requires that at least three of four successive points be located outside of the same one-sigma limit, would indicate that a signal does exist. Again, the evaluator would conclude that the difference scores changed after the financial aid allocation procedures were changed.

Insert Figure 3 about here

Discussion

In Total Quality Management, each element of education is viewed as a system or a subsystem. Continual improvement in the effectiveness of these systems is a goal under TQM. Thus, the function of evaluation under the TQM philosophy is to provide feedback on the effectiveness of changes made in the systems or subsystems.

An evaluator must use evaluation technique that is capable of separating a signal that a change in the system has had an impact on the outcome variable from the noise contained in that variable. One statistical quality control technique that educational evaluators can use to separate a signal from the noise contained in the outcome variable is an individual values chart. Educational evaluators who use individual values charts often use the three-sigma rule to detect a signal. This technique is effective at detecting large changes in the outcome variable. A change that is made in the educational system may produce a small or moderate change, but an educationally important change, in the outcome variable. The use of the three-sigma rule in conjunction with an individual values chart may not detect such a change.

By using the four Western Electric Zone Tests in conjunction with an individual values chart, an evaluator of an educational system or subsystem will increase the chance of detecting a true signal in the outcome variable. It should be noted that this increase in the chance of detecting a signal does not come without a price. The use of the four detection tests will increase the chance of identifying a change in the outcome variable as a signal when, in fact, the change

is simply a reflection of the normal variation in the outcome variable. That is, the evaluator has a greater chance of concluding that the change made in the educational system has had an impact on the outcome variable when the change had no such effect. As noted by Wheeler (1994), "each additional detection rule will increase the likelihood of false alarms" (p. 134). It is important for the evaluators to be cognizant of the strengths and weakness of the techniques that they employ.

References

Deming, W.E. (1986). Out of the Crisis (2nd ed.). Cambridge, MA: MIT Center for Advanced Engineering Study.

Juran, J.M. (Ed.). (1979). Quality Control Handbook. New York: McGraw-Hill.

Langford, D.P., & Cleary, B.A. (1995). Orchestrating learning with quality. Milwaukee, WI: ASQL Quality Press.

Newman, I., & Morsches, M. (1995, October). A systemic approach to evaluation: The difference between evaluation and assessment. Paper presented at the meeting of the Mid-Western Educational Research Association, Chicago, IL.

Newman, I., & Newman, C. (1993). The need for outcome rather than process evaluations. The Ohio Journal of Science, 93 (3), 68-71.

Shewhart, W.A. (1931). Economic control of quality of manufactured product. New York: D. Van Nostrand Company.

Statistical Quality Control Handbook. (1956). Western Electric Company. Indianapolis, IN: AT&T Technologies.

Wheeler, D.J. (1994). Advanced topics in statistical process control (2nd ed.) Knoxville, TN: SPC Press.

Wheeler, D.J. (1993). Understanding variation: The key to managing chaos. Knoxville, TN: SPC Press.

Table 1

Hypothetical Difference Values Used to Construct the Individual Values Chart

Year	Difference Scores[a]	Moving Range Values
1	-.7	—
2	-.1	.5
3	.2	.3
4	0	.2
5	-.4	.4
6	.1	.5
7	.3	.2
8	-.1	.4
9	-.6	.5
10	-.3	.3
11	-.6	.3
12	-.2	.4
Total	-2.4	4.1
	$\bar{x}$ = -.2	mR = .37

[a]A difference score is equal to the difference between the national mean ACT score and the mean ACT score of the students who enrolled at the university.

Table 2

Hypothetical Difference Scores for Scenario 1 and Scenario 2

Year	Difference Scores for Scenario 1[a]	Difference Scores for Scenario 2[a]
13	.5	.2
14	.6	.3
15		.2
16		.4

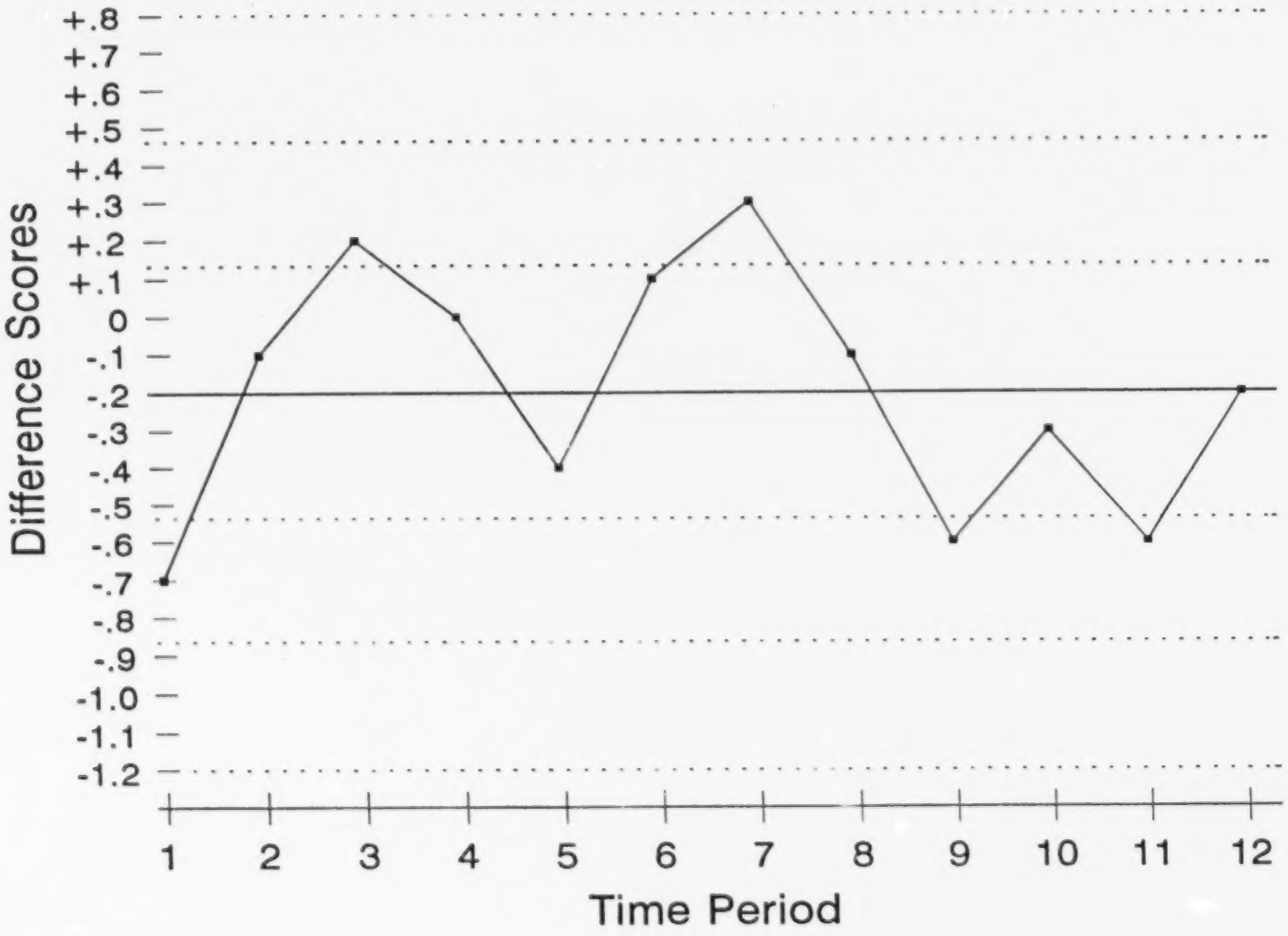
[a]A difference value is equal to the difference between the national mean ACT score and the mean ACT score of the students who enrolled at the university.

Figure Captions

Figure 1. Individual values chart based on the values prior to the changes in the financial aid program.

Figure 2. Individual values chart that includes the values for scenario 1.

Figure 3. Individual values chart that includes the values for scenario 2.


Difference Scores
+.8
+.7
+.6
+.5
+.4
+.3
+.2
+.1
0
-.1
-.2
-.3
-.4
-.5
-.6
-.7
-.8
-.9
-1.0
-1.1
-1.2
1
2
3
4
5
6
7
8
9
10
11
12
Time Period

I
TOP

Figure 1. Individual values chart based on the values prior to the changes in the financial aid program.

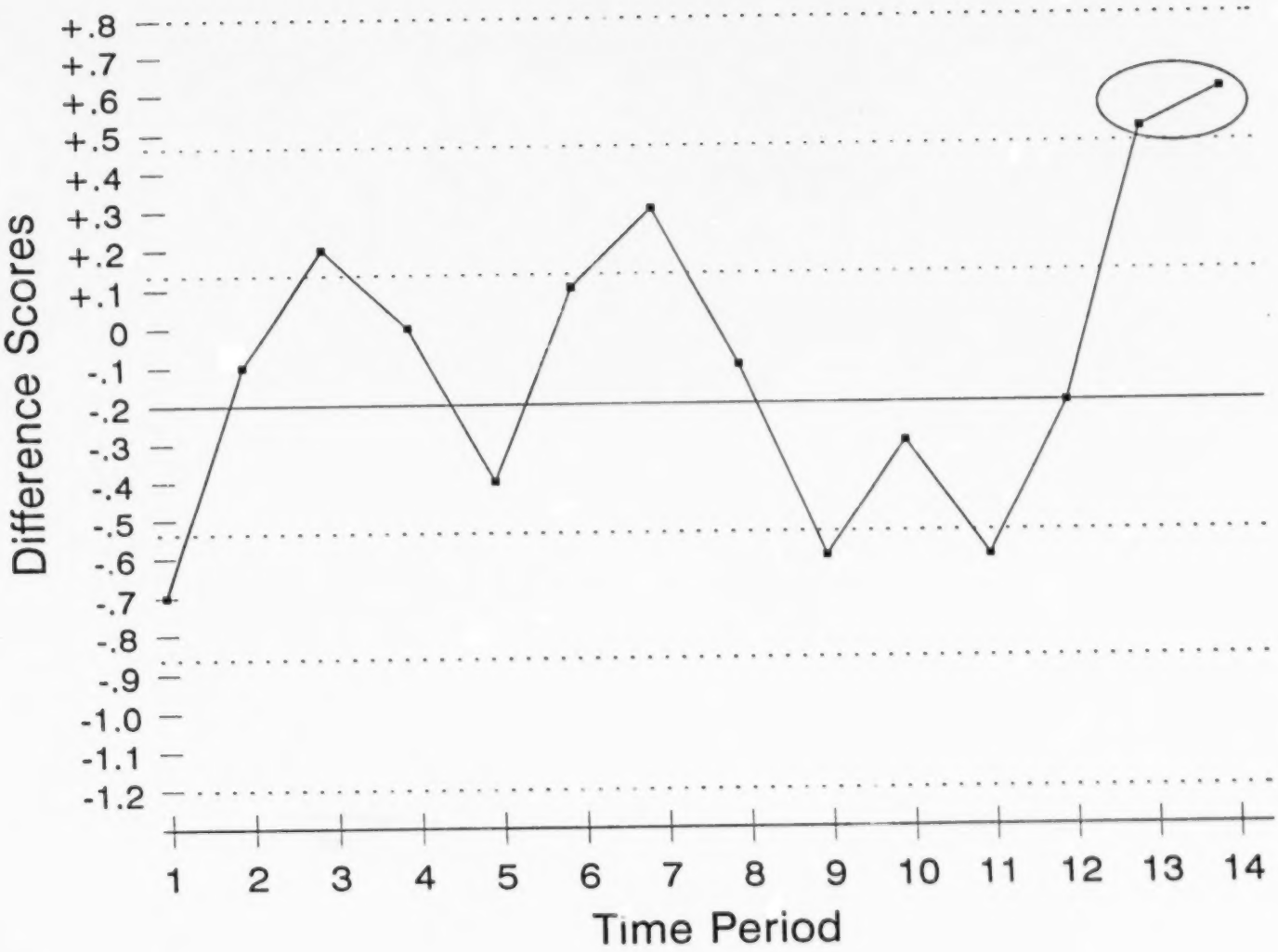
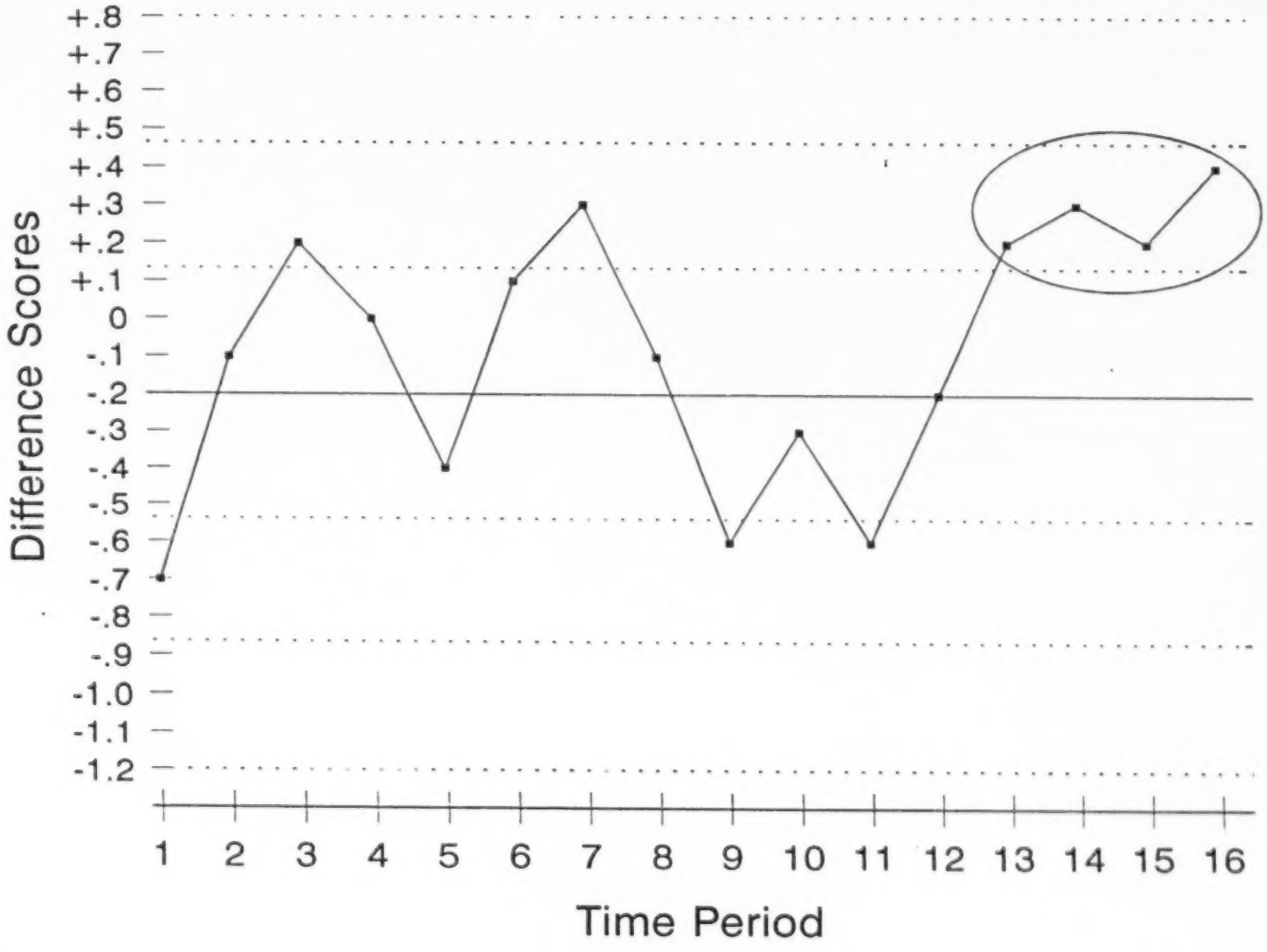

+.8
+.7
+.6
+.5
+.4
+.3
+.2
+.1
0
-.1
-.2
-.3
-.4
-.5
-.6
-.7
-.8
-.9
-1.0
-1.1
-1.2
Difference Scores
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Time Period

Figure 2. Individual values chart that includes the values for scenario 1.


+.8
+.7
+.6
+.5
+.4
+.3
+.2
+.1
0
-.1
-.2
-.3
-.4
-.5
-.6
-.7
-.8
-.9
-1.0
-1.1
-1.2
Difference Scores
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Time Period

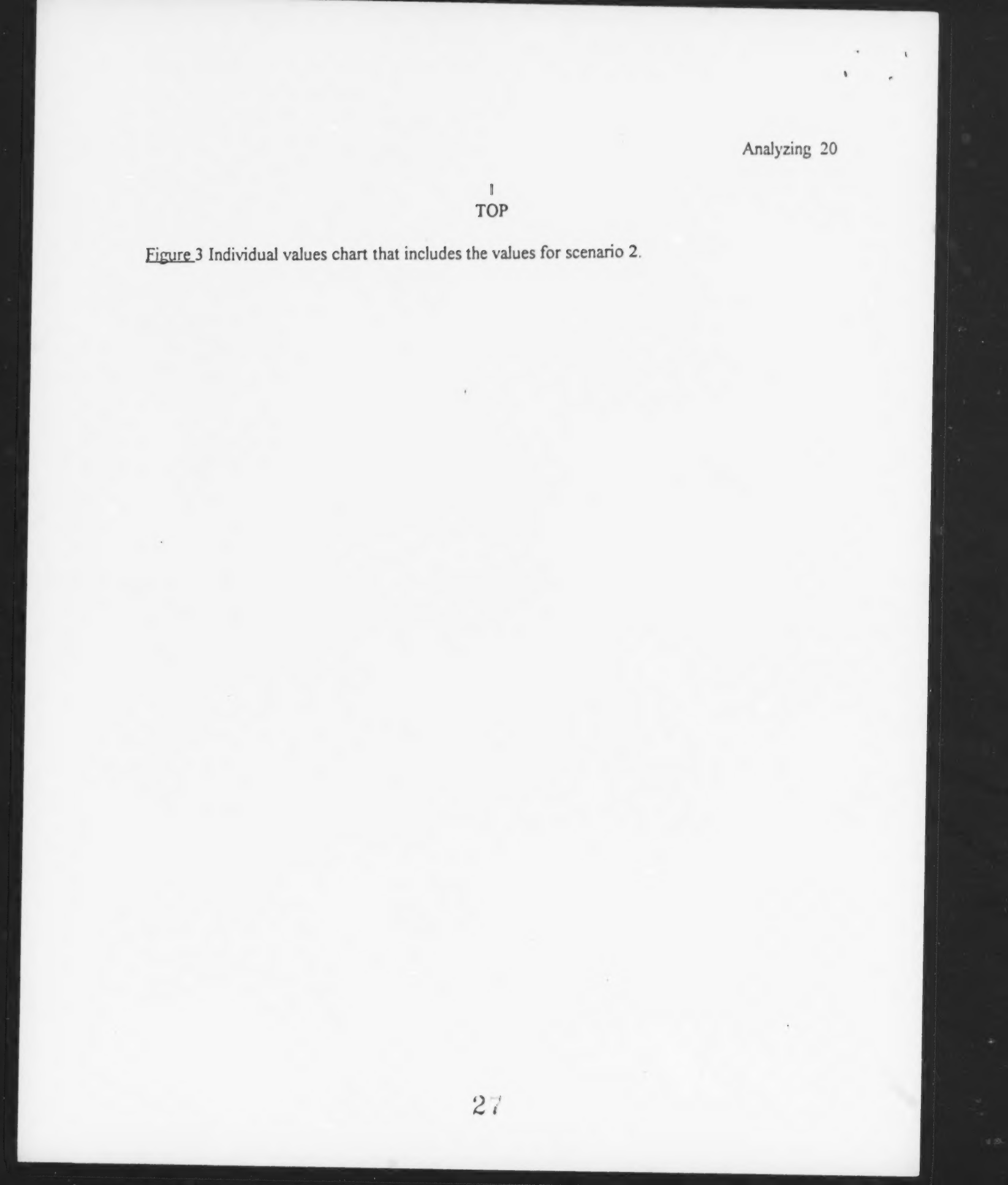


Figure 3 Individual values chart that includes the values for scenario 2.